

UC Riverside

UC Riverside Previously Published Works

Title

Classification and phylogenetic analyses of the Arabidopsis and tomato G-type lectin receptor kinases.

Permalink

<https://escholarship.org/uc/item/3pg58221>

Journal

BMC genomics, 19(1)

ISSN

1471-2164

Authors

Teixeira, Marcella A
Rajewski, Alex
He, Jiangman
et al.

Publication Date

2018-04-01

DOI

10.1186/s12864-018-4606-0

Peer reviewed

RESEARCH ARTICLE

Open Access



Classification and phylogenetic analyses of the Arabidopsis and tomato G-type lectin receptor kinases

Marcella A. Teixeira¹, Alex Rajewski², Jiangman He¹, Olenka G. Castaneda³, Amy Litt^{2,4} and Isgouhi Kaloshian^{1,4*} 

Abstract

Background: Pathogen perception by plants is mediated by plasma membrane-localized immune receptors that have varied extracellular domains. Lectin receptor kinases (LecRKs) are among these receptors and are subdivided into 3 classes, C-type LecRKs (C-LecRKs), L-type LecRKs (L-LecRKs) and G-type LecRKs (G-LecRKs). While C-LecRKs are represented by one or two members in all plant species investigated and have unknown functions, L-LecRKs have been characterized in a few plant species and have been shown to play roles in plant defense against pathogens. Whereas Arabidopsis G-LecRKs have been characterized, this family of LecRKs has not been studied in tomato.

Results: This investigation updates the current characterization of Arabidopsis G-LecRKs and characterizes the tomato G-LecRKs, using LecRKs from the monocot rice and the basal eudicot columbine to establish a basis for comparisons between the two core eudicots. Additionally, revisiting parameters established for Arabidopsis nomenclature for LecRKs is suggested for both Arabidopsis and tomato. Moreover, using phylogenetic analysis, we show the relationship among and between members of G-LecRKs from all three eudicot plant species. Furthermore, investigating presence of motifs in G-LecRKs we identified conserved motifs among members of G-LecRKs in tomato and Arabidopsis, with five present in at least 30 of the 38 Arabidopsis members and in at least 45 of the 73 tomato members.

Conclusions: This work characterized tomato G-LecRKs and added members to the currently characterized Arabidopsis G-LecRKs. Additionally, protein sequence analysis showed an expansion of this family in tomato as compared to Arabidopsis, and the existence of conserved common motifs in the two plant species as well as conserved species-specific motifs.

Keywords: G-LecRKs, Tomato, *Solanum lycopersicum*, *Arabidopsis thaliana*, Lectin receptor, Receptor kinase, *Aquilegia coerulea*, Columbine, Phylogenetic analysis

Background

In the constant war against pathogens, plants are equipped with a surveillance system that relies on pattern-recognition receptors (PRRs), proteins localized at the plasma membrane with ectodomains, that screen the environment for conserved microbial or pest- and damage-associated signals. In addition to the ectodomain, a subgroup of these PRRs has

intracellular kinase domains and are therefore known as receptor kinases (RKs). Plant RKs have undergone a recent expansion, with the *Arabidopsis thaliana* (Arabidopsis) genome encoding more than 600 RKs [1]. According to their ectodomains, RKs can be further classified into specific subgroups, such as leucine-rich repeat RKs (LRR-RKs) and lectin RKs (LecRKs). Receptor kinases are involved in several cellular processes, from adaptation to abiotic stresses to defense responses against pathogens and pests and interactions with microbial symbionts [2–14]. Several RKs and their cognate elicitor pairs have been described, mostly engaging the recognition of bacteria and fungi by plants [14–16].

* Correspondence: isgouhi.kaloshian@ucr.edu

¹Department of Nematology, University of California, Riverside, California, USA

⁴Institute for Integrative Genome Biology, University of California, Riverside, California, USA

Full list of author information is available at the end of the article



The best-characterized PRR-elicitor pair is the Arabidopsis LRR-RK FLS2 (FLAGELLIN SENSITIVE2) and the peptide flg22, consisting of a stretch of 22 amino acids of the N-terminal bacterial flagellin [17]. In addition to Arabidopsis, FLS2 orthologs have been identified in several plant species including tomato (*Solanum lycopersicum*), grapevine (*Vitis vinifera*), rice (*Oryza sativa*) and citrus (*Citrus paradisi*, *C. reticulata* and *Fortunella margarita*) [18–21]. Interestingly in tomato, a flagellin-derived peptide distinct from flg22, flgII-28, is perceived by the LRR-RK FLS3, and similar to FLS2, its perception and downstream signaling requires a second LRR-RK, BAK1 (BRASSINOSTEROID INSENSITIVE 1-ASSOCIATED KINASE 1) [22, 23]. Other receptor-ligand pairs include chitin perception by the lysin-motif RK (LysM-RKs) LYK5 and xylanase perception by the LRR-RK EIX2 [24, 25]. Although a co-receptor has not been characterized for xylanase perception, chitin perception requires participation of the LysM-RK CERK1 (CHITIN ELICITOR RECEPTOR KINASE 1) [26]. Interestingly, chitin perception in rice is mediated by the LysM-receptor like protein CEBiP (CHITIN ELICITOR BINDING PROTEIN), which lacks a kinase domain and relies on its co-receptor CERK1 for kinase signaling of chitin perception [27].

The LecRKs are a second type of receptors known for their role in binding various carbohydrates [9]. Based on their ectodomains, LecRKs can be classified into C-type, L-type or G-type (Fig. 1). The C-type (calcium-dependent) LecRKs (C-LecRK) contain the C-type motif that is commonly found in several proteins from mammals, and these proteins have been shown to have a role in innate immunity [10, 28, 29]. Interestingly, in plants this LecRK group is represented by only a single gene in Arabidopsis, rice, and tomato, and two genes in wheat (*Triticum aestivum*) [28, 30, 31].

The L-type (legume-like) LecRKs (L-LecRKs) consist of large number of family members and have well-characterized roles in plant defense. Investigations in Arabidopsis, rice, tomato, *Nicotiana benthamiana* and wheat revealed 45, 72, 22, 37, 84 members of this family, respectively [28, 30, 31]. Several reports link genes of this family to defense against pathogens; for example, AtLecRK-I.9 against the bacterial pathogen *Pseudomonas syringae* pv. *tomato* (*Pst*) [32], AtLecRK-IX.I and LecRK-IX.2 against the pathogenic oomycetes *Phytophthora brassicae* and *P. capsici* [33], AtLecRK-I.9 against *P. infestans* [34], and AtLecRK-VI.2 against the pathogenic bacteria *P. syringae* and *Pectobacterium carotovorum* [13, 35]. Additionally, L-LecRKs have been implicated in perception of the danger molecule, the extracellular ATP, by the AtLecRK-I.9 [36, 37].

The G-type LecRKs (G-LecRKs) are proteins with an ectodomain that resembles the *Galanthus nivalis* agglutinin (GNA) mannose-binding motif [10, 38]. Since the discovery of the first G-LecRK, GNA, in 1987, this group has been renamed and refined several times [39]. The specificity of GNA for binding mannose differentiated it from the mannose/glucose specificity of L-LecRKs and led other proteins of this group to be generically called alpha-D-mannose-specific lectins [39, 40]. Soon after the discovery of GNA, several studies targeting other members of Amaryllidaceae, along with Orchidaceae and Alliaceae discovered more mannose-specific lectins [41, 42]. Based on the apparently narrow taxonomic occurrence of the proteins at the time, the group was then renamed monocot mannose-binding lectins. Early amino acid sequence-based analyses also began to suggest that these proteins might be encoded by a large multigene family [43]. Subsequently, similar proteins were discovered in both liverwort (*Marchantia polymorpha*) and yew (*Taxus media*) necessitating another renaming to

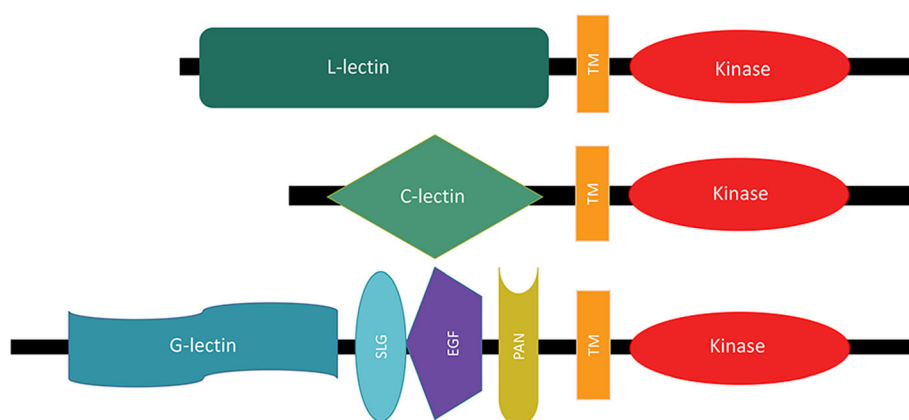


Fig. 1 Domains of lectin receptor kinases (LecRKs). G-lectin, C-lectin and L-lectin are the motifs in the ectodomains of G-type, C-type, and L-type LecRKs, respectively

the more general GNA-related lectins [44, 45]. Because GNA was first isolated from *Galanthus nivalis* bulbs and other G-LecRKs were present in high amounts in bulbs of other species, this group has also been called Bulb- or B-type lectins. This name has been widely adopted because it is used to identify this domain in both the PROSITE and Pfam databases (PS50927 and PF01453, respectively). Despite the widespread use of the “B-type” descriptor in databases and genome annotations, GNA-related or G-type lectin is currently the preferred term for these proteins.

Previous investigations identified 32 members of G-LecRKs in Arabidopsis, 100 in rice and 177 in wheat [28, 31]. The best-known members of this group are the S-locus (S-locus glycoprotein/SLG containing) RKs, known for their role in self-incompatibility in flowering plants [46, 47]. Besides the G-type lectin and the kinase domains, G-LecRKs can have additional domains such as a cysteine-rich domain (Fig. 1), known as the epidermal growth factor (EGF) domain, which is thought to play a role in disulfide bond formation [48]. Additionally, family members may contain the plasminogen-apple-nematode (PAN) motif, which likely has a role in protein-protein or protein-carbohydrate interactions (Fig. 1) [49].

Typically, members of large families do not have consistent nomenclature, as frequently not all members are identified at the same time, and gains and losses in different species make identification of orthologous genes and gene groups difficult. While the Arabidopsis L-LecRK family members have a clear systematic nomenclature based on chromosome location and amino acid and nucleotide identity [34], currently members of the G-LecRKs do not have such nomenclature. Similarly, although L-LecRKs have been characterized in different plant species [28, 30, 31], tomato G-LecRKs have not been described to date. Given the role of these proteins in defense, an understanding of their diversity in tomato is critical to improving crop resistance to diseases and pests.

For this work, we searched the genomes of Arabidopsis and tomato to identify and characterize G-LecRKs, and performed phylogenetic analyses on the aligned sequences, using C-LecRKs and L-LecRKs as outgroups. To draw inferences regarding expansion vs loss within gene clades, we included G-LecRKs from the monocot rice and the basal eudicot columbine (*Aquilegia coerulea*), which diverged prior to the origin of the core eudicot clade that includes both Arabidopsis and tomato and thus serves to polarize the evolutionary trends. The analyses allowed identification of incorrect gene annotations in genome databases, and evaluation of sequence similarity between G-LecRKs and identification of

instances of gene clade expansion or gene loss in Arabidopsis and tomato. Based on this investigation, we suggest a nomenclature for members of this gene family from both Arabidopsis and tomato.

Methods

Database searches, protein domain and genome organization

To identify Arabidopsis (*Arabidopsis thaliana*) G-LecRKs a first search was performed using the lectin domain of At1g61550 as the query followed by the lectin domain of At1g61400, At2g19130, At4g21390 and At5g60900 for a second search in The Arabidopsis Information Resource (TAIR) (<http://arabidopsis.org>) website. Results with e-value < 0.1 were considered G-LecRKs candidates. The localization of G-LecRKs on the Arabidopsis genome was visualized using the chromosomal map tool from TAIR (<http://arabidopsis.org/jsp/ChromosomeMap/tool.jsp>). Arabidopsis predicted kinase domain sequences were aligned using ClustalW and the alignment was manually checked to identify the kinase subdomains using AtLecRK-VI.2 as a reference [13, 30].

To identify tomato (*Solanum lycopersicum*) G-LecRKs, the At1g61550 lectin domain was used as the query in the Sol Genomics Network (SGN) (<https://solgenomics.net>) and at the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/>) websites. Results with e-value < 0.1 were considered G-LecRKs candidates. After the initial search using the At1g61550 G-lectin domain, a second search was performed in NCBI and both searches were cross analyzed to compile a list of all possible G-LecRK candidates. Tomato G-LecRKs sequences were mapped according to Shearer et al. [50]. To map tomato sequences onto chromosomes, gene position information was acquired from Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>) and JBrowse at SGN. Mapping of G-LecRKs onto tomato chromosomes was performed manually using NCBI Map Viewer (<https://www.ncbi.nlm.nih.gov/genome/gdv/>). Tomato predicted kinase domain sequences were aligned with the kinase domain of Solyc03g006720 using ClustalW and the kinase subdomains were manually checked to identify the kinase subdomains.

To identify columbine (*Aquilegia coerulea*) G-LecRKs, the same Arabidopsis gene (At1g61550) was used in a BLASTp search of the genome sequence hosted by Phytozome. Three hits from this search were then chosen for another round of searching with BLASTp. Only BLAST hits with an e-value < 0.1 were retained. A search was also conducted using the keyword “lectin”. The resulting sequences were passed to IPRscan on the University of California at Riverside (UCR) High-Performance Computing Center (HPCC) cluster for protein domain identification using Pfam and Prosite [51].

For G-LecRK identification in rice (*Oryza sativa*), in addition to the 87 proteins listed in Vaid et al. [28], protein sequences annotated with the Pfam code PF01453 were retrieved from the Rice Genome Annotation Project (<http://rice.plantbiology.msu.edu/>). All sequences were passed to IPRscan using the same parameters as for columbine.

Phylogenetic analysis

Full-length protein sequences were downloaded as above from TAIR (Arabidopsis), Sol Genomics Network (tomato), Phytozome (columbine), or the Rice Genome Annotation Project (rice) and aligned using MUSCLE on the UCR HPCC cluster [52]. Eight outgroup protein sequences representing one C-LecRK and L-LecRK each from Arabidopsis, tomato, rice and columbine were included in the alignments. These sequences were from the single Arabidopsis C-LecRK gene, At1g52310, and its top BLASTp hits from tomato (Solyc02g068370.2), rice (Os01g010400.1), and columbine (Aqcoe2G393700); and an Arabidopsis L-LecRK (At1g52310) and its top BLASTp hits from tomato (Solyc02g068370.2), rice (Os04g0531500.1) and columbine (Aqcoe2G393700). The initial sequence alignment was generated using data from Arabidopsis, tomato, columbine, and rice. Subsequently, a three-species alignment was generated using only sequences from Arabidopsis, tomato, and columbine.

The JTT model of protein sequence evolution was determined to best fit the data based on Akaike Information Criteria using the web server version of SMS, and phylogenetic trees were subsequently constructed from both the four-species and three-species amino acid alignments using this evolutionary model in RAxML v8 on the UCR HPCC cluster with 1000 bootstraps [53–55]. A collapsed tree was also constructed from the three- and four- species trees in TreeGraph 2 by collapsing any node with < 70% bootstrap support into a polytomy [56].

Domain and motif identification

Protein domains were investigated using multiple online programs including, InterPro (<https://www.ebi.ac.uk/interpro/>) and TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>). Investigation of conserved motifs in the ectodomains of Arabidopsis and tomato G-LecRKs was performed using the default settings at MEME (Multiple EM for Motif Elicitation) Suite 4.11.2 (<http://meme-suite.org/tools/meme>) [57].

Subcellular localization prediction

Multiple protein subcellular localization tools were used to localize the Arabidopsis and tomato G-LecRKs. Arabidopsis gene identifiers were used to query “The SUBcellular localization database for Arabidopsis proteins”, SUBA3

(<http://suba3.plantenergy.uwa.edu.au/>) [58, 59]. Additionally, amino acid sequences of both Arabidopsis and tomato G-LecRKs were analyzed using TargetP 1.1 Server (<http://www.cbs.dtu.dk/services/TargetP/>) and “subCELLular LOcalization predictor” CELLO v.2.5 (<http://cello.life.nctu.edu.tw/>) [60, 61].

Results

Annotation of Arabidopsis G-LecRKs

To identify the Arabidopsis G-LecRKs, a BLASTp analysis [62] was performed at the TAIR website using the region comprising the predicted G-type lectin domain [63], amino acids 24–170 from At1g61550. The search resulted in 44 sequences. From these sequences, four (At1g61400, At2g19130, At4g21390 and At5g60900) were chosen for use as new queries to fish out additional candidates. These analyses resulted in a total of 49 proteins with a G-lectin domain (Table 1). The majority of these proteins had SLG (75%) and PAN (82%) domains but only 10 had an EGF domain. Of these 49 sequences, 38 proteins had also kinase domains and were considered for further analyses (Table 1).

Previous characterization of Arabidopsis G-LecRKs included 31 sequences [28], all also identified in our search. This same study also identified a protein, At1g61460, that was not detected in our search. To confirm the identity of this protein, it was used in domain search using InterPro. Domain predictions showed that At1g61460 has SLG, PAN, transmembrane (TM) and kinase domains, but not a lectin domain. Because of the absence of the lectin domain, this protein was not considered a G-LecRK and was not used in further analyses.

Annotation of tomato G-LecRKs

Using the same strategy used to retrieve the Arabidopsis G-LecRKs, the tomato genome was queried for G-type lectin homologs using the lectin domain of At1g61550. Two databases, SGN and NCBI, were searched. The search against SGN resulted in 21 sequences. The search against NCBI resulted in numerous hits, including a number of different isoforms of variable lengths of the same protein. The combined results from these two searches yielded 88 distinct sequences with a G-type lectin domain (Table 2). To assure a comprehensive search, three random tomato G-type lectins were chosen to query the tomato genome again using their predicted G-lectin domain. This resulted in two additional candidates, Solyc07g053220 and Solyc05g008310.

Three of the identified G-type lectin-containing sequences were misannotated. Solyc03g006730, Solyc04g008400, and Solyc07g055640 each contained two G-LecRKs in tandem and were therefore each split into two (Solyc03g006730.A and Solyc03g006730.B; Solyc04g008400.A and Solyc04g008400.B; and Solyc07g055640.

Table 1 Domains of Arabidopsis genes encoding G-Lectins

	Locus	^a SLG	EGF	PAN	TM	Kinase
1	AT1G11340	^b x	x	x	x	x
2	AT1G11410	x	x	x	x	x
3	AT1G61360	x	x	x	x	x
4	AT1G61380	x	x	x	x	x
5	AT1G61390	x	x	x	x	x
6	AT1G61550	x	x	x	x	x
7	AT1G61610	x	x	x	x	x
8	AT2G19130	x	x	x	x	x
9	AT4G27290	x	x	x	x	x
10	AT4G03230	x	x	x	x	x
11	AT1G11280	x	–	x	x	x
12	AT1G11300	x	–	x	x	x
13	AT1G11305	x	–	x	x	x
14	AT1G11330	x	–	x	x	x
15	AT1G11350	x	–	x	x	x
16	AT1G61370	x	–	x	x	x
17	AT1G61400	x	–	x	x	x
18	AT1G61420	x	–	x	x	x
19	AT1G61430	x	–	x	x	x
20	AT1G61440	x	–	x	x	x
21	AT1G61480	x	–	x	x	x
22	AT1G61490	x	–	x	x	x
23	AT1G61500	x	–	x	x	x
24	AT1G65790	x	–	x	x	x
25	AT1G65800	x	–	x	x	x
26	AT2G41890	x	–	x	x	x
27	AT4G11900	x	–	x	x	x
28	AT4G21380	x	–	x	x	x
29	AT4G21390	x	–	x	x	x
30	AT4G27300	x	–	x	x	x
31	AT1G34300	x	–	–	x	x
32	AT4G00340	x	–	–	x	x
33	AT5G24080	x	–	–	x	x ^c
34	AT3G16030	–	–	x	x	x
35	AT4G32300	–	–	–	x	x
36	AT5G35370	–	–	–	x	x
37	AT1G67520	–	–	x	–	x ^c
38	AT5G60900	–	–	x	–	x
39	AT5G03700.1	x	–	x	x	–
40	AT1G78830	–	–	x	x	–
41	AT3G12000	x	–	x	x	–
42	AT3G51710.1	x	–	x	x	–
43	AT1G16905	–	–	–	x	–
44	AT5G18470	–	–	–	x	–

Table 1 Domains of Arabidopsis genes encoding G-Lectins (Continued)

	Locus	^a SLG	EGF	PAN	TM	Kinase
45	AT1G78820	–	–	x	–	–
46	AT1G78850	–	–	x	–	–
47	AT1G78860	–	–	x	–	–
48	AT2G01780	–	–	–	–	–
49	AT5G39370	–	–	–	–	–

^aSLG S-locus glycoprotein, EGF epidermal growth factor, PAN plasminogen apple nematode, TM transmembrane

^b"x" denotes presence and "–" denotes absence of a domain

^cIncomplete kinase domain

A and Solyc07g055640.B) (Table 2). Thus, a total of 93 tomato sequences were identified with G-type lectin domains. The majority (72%) of these tomato sequences had an SLG domain and about half (63.5%) had a PAN domain. However, similar to Arabidopsis, the great majority lacked the EGF domain, with only seven proteins containing this domain (Table 2). Of the 93 tomato sequences, 73 proteins had both G-type lectin and kinase domains and were considered G-LecRKs for further analyses.

Annotation of columbine G-LecRKs

For columbine G-LecRK identification, the same Arabidopsis G-LecRK sequence, At1g61550, was used as a query sequence for a BLASTp search of the genome sequence. Three hits from this search were then chosen for another round of searching with BLASTp. Taking advantage of the functional genome annotations available, a keyword search was also conducted of the columbine genome using the keyword "lectin". After merging duplicates, this yielded 166 unique protein sequences. Of these, two could not be annotated at all, while 43 others lacked a G-type lectin domain, kinase domain, or both (Additional file 1). Of the 121 columbine sequences, 59 proteins had both G-type lectin and kinase domains and were used in the phylogenetic analysis.

Annotation of rice G-LecRKs

For G-LecRK identification in rice, the 87 previously published proteins were all included [28]. In addition, protein sequences annotated with the Pfam code PF01453 (B-type lectin, synonymous with G-type lectin) were retrieved from the genome sequence. This generated 143 sequences, including the 87 previously reported, but also including different isoforms of some proteins. In the case of isoforms, only the longest was retained. Additionally, LOC_Os09g37840.1 appeared to be a misannotation containing three G-LecRKs in tandem. This locus was split into three sequences (LOC_Os09g37840.1.A., B, and C). This resulted in 145 protein sequences with G-type lectin domain (Additional file 2).

Table 2 Domains of tomato genes encoding G-Lectins

	Locus	^a SLG	EGF	PAN	TM	Kinase
1	Solyc02g079640	^b x	x	x	x	x
2	Solyc04g008400.A	x	x	x	x	x
3	Solyc04g008400.B	x	x	x	x	x
4	Solyc04g058110	x	x	x	x	x
5	Solyc07g063770	x	x	x	x	x
6	Solyc10g006710	x	x	x	x	x
7	Solyc11g005630	x	x	x	x	x
8	Solyc01g094830	x	—	x	x	x
9	Solyc02g030300	x	—	x	x	x
10	Solyc02g079530	x	—	x	x	x
11	Solyc02g079540	x	—	x	x	x
12	Solyc02g079550	x	—	x	x	x
13	Solyc02g079570	x	—	x	x	x
14	Solyc02g079590	x	—	x	x	x
15	Solyc02g079620	x	—	x	x	x
16	Solyc02g079630	x	—	x	x	x
17	Solyc02g079710	x	—	x	x	x
18	Solyc03g006720	x	—	x	x	x
19	Solyc03g006730.A	x	—	x	x	x
20	Solyc03g006730.B	x	—	x	x	x
21	Solyc03g006770	x	—	x	x	x
22	Solyc03g006780	x	—	x	x	x
23	Solyc03g063650	x	—	x	x	x
24	Solyc04g008370	x	—	x	x	x
25	Solyc04g077270	x	—	x	x	x
26	Solyc04g077280	x	—	x	x	x
27	Solyc04g077300	x	—	x	x	x
28	Solyc04g077340	x	—	x	x	x
29	Solyc04g077360	x	—	x	x	x
30	Solyc04g077370	x	—	x	x	x
31	Solyc04g077390	x	—	x	x	x
32	Solyc04g078410	x	—	x	x	x
33	Solyc05g008310	x	—	x	x	x
34	Solyc07g053080	x	—	x	x	x
35	Solyc07g053120	x	—	x	x	x
36	Solyc07g053130	x	—	x	x	x
37	Solyc07g053220	x	—	x	x	x
38	Solyc07g063700	x	—	x	x	x
39	Solyc07g063710	x	—	x	x	x
40	Solyc07g063720	x	—	x	x	x
41	Solyc07g063730	x	—	x	x	x
42	Solyc07g063750	x	—	x	x	x
43	Solyc07g063780	x	—	x	x	x
44	Solyc07g063800	x	—	x	x	x

Table 2 Domains of tomato genes encoding G-Lectins
(Continued)

	Locus	^a SLG	EGF	PAN	TM	Kinase
45	Solyc09g011330	x	—	x	x	x
46	Solyc10g005440	x	—	x	x	x
47	Solyc10g006720	x	—	x	x	x
48	Solyc12g005290	x	—	x	x	x
49	Solyc01g006520	x	—	—	x	x
50	Solyc03g005130	x	—	—	x	x
51	Solyc03g007790	x	—	—	x	x
52	Solyc03g078360	x	—	—	x	x
53	Solyc03g078370	x	—	—	x	x
54	Solyc06g036470	x	—	—	x	x
55	Solyc09g075910	x	—	—	x	x
56	Solyc09g075920	x	—	—	x	x
57	Solyc11g013880	x	—	—	x	x
58	Solyc02g072070	—	—	x	x	x
59	Solyc03g120110	—	—	x	x	x
60	Solyc08g076050	—	—	x	x	x
61	Solyc08g076060	—	—	x	x	x
62	Solyc12g006840	—	—	x	x	x
63	Solyc01g006530	—	—	—	x	x
64	Solyc04g015460	—	—	—	x	x
65	Solyc04g077380	—	—	—	x	x
66	Solyc07g055650	—	—	—	x	x
67	Solyc08g059730	—	—	—	x	x
68	Solyc07g063820	x	—	x	—	x
69	Solyc07g063810	—	—	x	—	x
70	Solyc07g055630	—	—	—	—	x
71	Solyc07g055640.A	—	—	—	—	x
72	Solyc07g055640.B	—	—	—	—	x
73	Solyc08g076070	—	—	—	—	x
74	Solyc04g077310	x	—	—	x	—
75	Solyc07g053090	x	—	x	x	—
76	Solyc09g009150	x	—	x	x	—
77	Solyc09g018490	x	—	—	x	—
78	Solyc02g076830	x	—	x	—	—
79	Solyc10g006690	x	—	x	—	—
80	Solyc04g077320	x	—	—	—	—
81	Solyc07g009440	x	—	—	—	—
82	Solyc07g055690	x	—	—	—	—
83	Solyc01g014510	—	—	—	—	—
84	Solyc01g014540	—	—	—	—	—
85	Solyc01g014560	—	—	—	—	—
86	Solyc01g014640	—	—	—	—	—
87	Solyc01g014700	—	—	—	—	—

Table 2 Domains of tomato genes encoding G-Lectins
(Continued)

	Locus	^a SLG	EGF	PAN	TM	Kinase
88	Solyc02g030340	–	–	–	–	–
89	Solyc02g030380	–	–	–	–	–
90	Solyc02g078730	–	–	–	–	–
91	Solyc07g009410	–	–	–	–	–
92	Solyc07g062480	–	–	–	–	–
93	Solyc07g062490	–	–	–	–	–

^aSLG S-locus glycoprotein, EGF epidermal growth factor, PAN plasminogen apple nematode, TM transmembrane

^b"x" denotes presence and "–" denotes absence of a domain

Of these 145, 122 rice sequences had also a kinase domain and were used in the phylogenetic analysis.

Phylogenetic analysis

Alignment of the 38 Arabidopsis, 73 tomato, 59 columbine, and 122 rice putative G-LecRK proteins was used to construct a phylogenetic tree with 1000 bootstrap replicates using RaxML [55]. Orthologous copies of a single C- and L-LecRK from Arabidopsis, tomato, rice, and columbine were used as outgroups to root the tree. In this four-species analysis, the C-LecRKs form a sister-clade to a large clade that includes all the other genes, including the L-LecRKs (Additional file 3). The clade containing all of the putative G-LecRKs and the intended outgroup L-LecRKs is further divided into two large clades, although with weak support. The L-LecRKs are included as the first branch in one of these two clades, suggesting that in a phylogenetic context, they should be considered G-LecRKs. In general, support in this tree for deep nodes is weak, with much stronger support towards the tips, suggesting rapid diversification of this gene group, and making interpretation of clade relationships difficult. In addition, the large number of rice genes largely cluster into clades that are nearly or entirely rice-specific (Additional file 3). This indicates extensive expansion of the G-LecRKs independently in rice relative to the eudicots, and makes interpretation of the relationships between Arabidopsis and tomato genes more difficult. For that reason, we performed a second phylogenetic analysis including only Arabidopsis, tomato, and columbine.

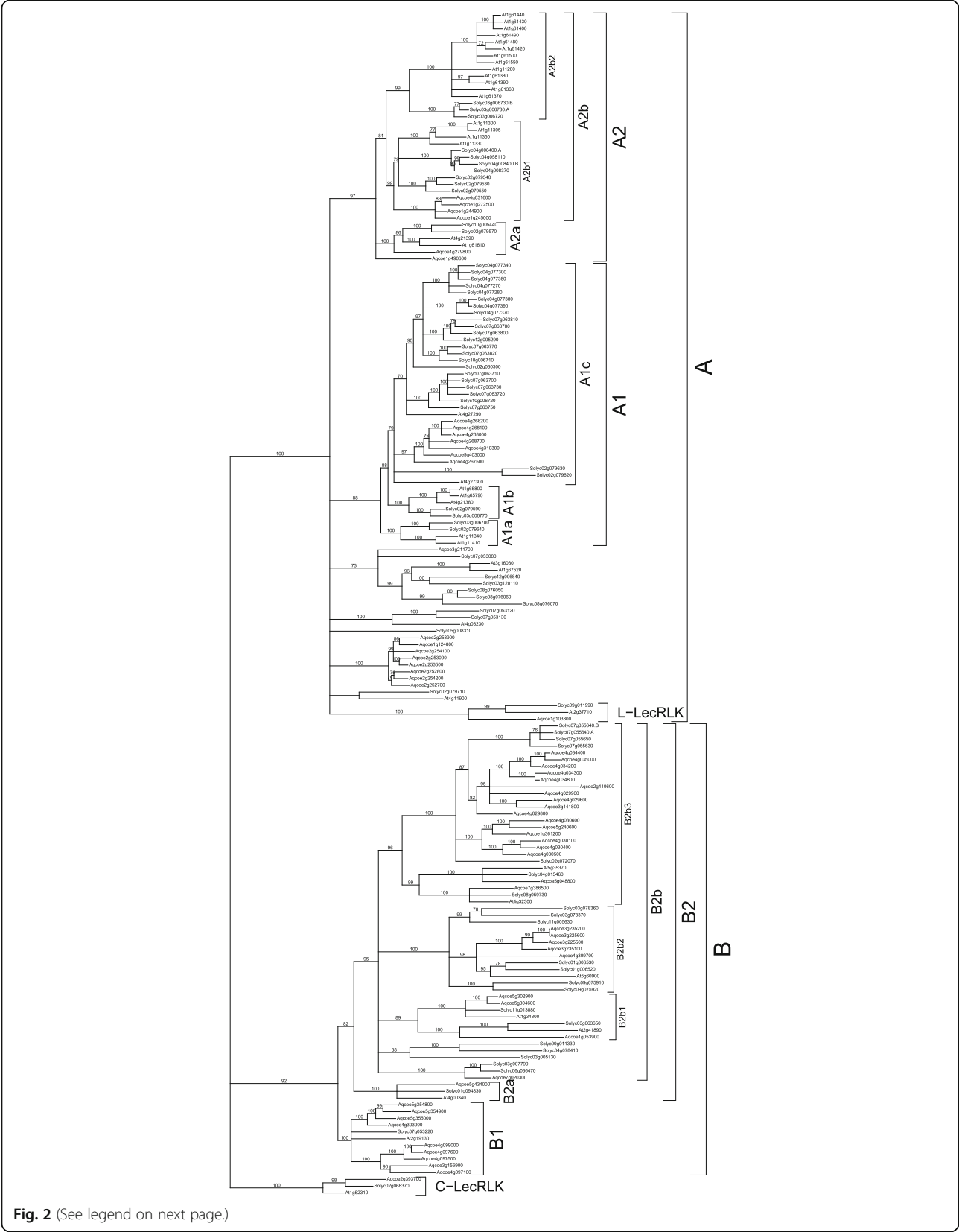
The results of this three-species analysis (Additional file 4) mirror those of the four-species analysis, with the C-LecRK outgroup as sister to a large clade that includes both the L- and G-LecRKs. As in the first analysis, this clade is further subdivided into two major clades, one of which (Clade A) includes the L-LecRKs, again suggesting they are not a group distinct from the G-LecRKs. Also consistent with the first analysis, support for deep nodes is weak. For instance, the first branch of clade A (Additional file 4) consists of the three L-LecRK genes that were intended to serve as

outgroup with the C-LecRKs. The next branch, with 30% support, includes only one gene each from Arabidopsis and tomato, and the following branch, with 24% support, includes 8 genes from columbine only. This implies that a columbine gene has been lost from the first clade, and that Arabidopsis and tomato genes have been lost from the second. With such low support, however, it is not possible to rule out a topology in which all these genes are members of a single clade, with one Arabidopsis, one tomato, and 8 columbine genes, indicating diversification in columbine rather than multiple losses.

Because of the weak support at many key nodes, for analysis of clade relationships and membership, we used a tree in which all nodes with less than 70% support were collapsed into polytomies (Fig. 2). Although the resulting topology includes fewer resolved relationships, those represented are more robustly supported in the data, providing a stronger basis for exploring expansions and losses in Arabidopsis and tomato. The two large clades remain, although their relationship to the outgroup is unresolved. This suggests an early duplication, prior to the divergence of monocots, producing two independently diverging clades of G-LecRKs. Available data do not show a clear distinction between these groups in function or expression; members of both groups show response (functional or regulatory) to various biotic and abiotic stresses [47, 64–69]. Several members of clade A (Fig. 2) are implicated in growth and development processes, which have so far not been reported for clade B members, however relatively few clade B members have been characterized at any level [70, 71].

A glance at the phylogenetic tree shows that within clades, there has been dramatically differential expansion of specific G-LecRK lineages in specific species. For the most part, Arabidopsis, tomato, and columbine genes form distinct species-specific groups within larger clades, suggesting that there have been multiple duplications leading to a large number of gene clades, but that after these duplications, there has been separate diversification of gene groups within each species. Expression and functional data are currently available only for Arabidopsis genes, and the diversity of functions, even within a single clade, suggests that this is a rapidly evolving group [66, 72–80]. This makes prediction of function in tomato difficult in most cases.

Clade A1, with 88% support, includes a total of 7 columbine, 7 Arabidopsis, and 28 tomato sequences. This clade is further subdivided into two successive lineages with multiple Arabidopsis and tomato genes but no columbine gene, suggesting loss of columbine genes and additional duplications separately within both Arabidopsis and tomato. The first clade (A1a) includes the Arabidopsis gene At1gl1340.1, the expression of which has been



(See figure on previous page.)

Fig. 2 Maximum likelihood tree of amino acid sequences of G-LecRK sequences from Arabidopsis, tomato, and columbine, with L-LecRK and C-LecRK sequences from each species as outgroup (but see text regarding L-LecRK sequence placement). Bootstrap support from 1000 replicates is shown above nodes. Nodes with < 70% bootstrap support were collapsed into polytomies. Brackets on the right indicate hierarchical clade names (or outgroups) as defined in the text

shown to be suppressed by mevalonic acid [73]. The second (A1b) includes the S-locus/ARK3 gene At4g21380.1, involved in self incompatibility, which is sister to the sister-gene pair At1g65790.1, upregulated by *Fusarium*, salt, and flg22, and At1g65800.1 (SD1-6/ARK2), involved in lateral root formation [78, 79, 81, 82]. Clade A1a is sister to a large clade (A1c) with poor internal resolution that includes 7 columbine, 2 Arabidopsis, and 24 tomato genes, indicating dramatic relative expansion within tomato relative to the other two species. Of the two Arabidopsis genes found in this clade, At4g27300.1 is upregulated by osmotic stress and trehalose, and downregulated by sucrose, whereas At4g27290.1 is induced by cold [72, 83]. The wide variation in reported functions for these Arabidopsis genes, and the lack of data for the other species, makes it difficult to generalize regarding the function of genes of individual clades and therefore to extrapolate and predict the functions of the tomato genes.

Clade A includes one other large clade with 97% support (A2) that is further subdivided into two clades with one columbine gene (Aqcoelg490600) that does not fall into either. Clade A2a, with 100% support, includes 5 genes: one columbine and two each Arabidopsis and tomato. Because the two tomato genes form a sister-pair, as do the two Arabidopsis genes, this indicates independent duplications in those two species. At1g61610.1 has been shown to be upregulated by flg22, bacterial lipopolysaccharide (LPS), and *Pst* strain DC3000, suggesting a role in defense against bacterial pathogens [69, 84]. Its paralog At4g21390.1 is also implicated in defense and is upregulated in the presence of fungal elicitors [72]. Although it is clear the function of these genes is highly labile, overall it suggests a role in defense for this clade.

Clade A2b, with 81% support, includes two subclades, both with 99% support. Both clades include sequences from both Arabidopsis and tomato, but only one (A2b1) also includes columbine, indicating a duplication that produced the two clades but loss of the columbine gene from one of the clades (A2b2). Both of these sister-clades have experienced expansion in both tomato and Arabidopsis, but A2b1 has a somewhat greater number in tomato (7 vs 4 in Arabidopsis), whereas A2b2 has the reverse, with substantially more in Arabidopsis (13 vs 3 in tomato). The presence of distinct Arabidopsis and tomato genes again suggests, in these two clades, independent expansion in the two species. Clade A2b1

includes Arabidopsis EGM1 (ENHANCED GROWTH ON MANNITOL1; At1g11300) and EGM2 (AT1g11305), paralogs implicated in plant shoot growth and mannitol stress, as well as CBRLK1 (CALMODULIN-BINDING RECEPTOR-LIKE CYTOPLASMIC KINASE1; At1g11350), a negative regulator of immunity against *Pst*, and At1g1130.2, which likely also plays a role in bacterial immunity [65, 70, 72]. The topology of the clade suggests that the role for shoot growth is derived. Expression of the Arabidopsis genes in clade A2b2 have been shown to respond to a wide variety of factors and, as with other clades, to play roles in plant development and defense responses [71, 75, 76, 84, 85]. Among these is the LPS receptor LORE (LIPOPOLYSACCHARIDE-SPECIFIC REDUCED ELICITATION/ SD1-29; At1g61380), known to mediate LPS sensitivity in Brassicaceae [66]. This LPS sensitivity does not appear to be in solanaceae [66].

Patterns of diversification are more varied and complex in clade B, and it includes only 7 genes from Arabidopsis, along with 23 from tomato and 37 from columbine. This clade is divided into two subclades, one of which (B1, 100% support) shows diversification in columbine, with 9 genes compared to one each in Arabidopsis and tomato. Although this Arabidopsis gene (At2g19130) has not been characterized, its ortholog in rice (*OsSIK2*; Os07g0186200) is implicated in salt and drought response [86]. The other clade, B2 (82% support), is further subdivided into one small clade that includes only one gene from each species, and a large clade with multiple subclades. This small clade B2a (100% support), one of the very few clades in the tree that has only one gene per species, has no evidence of diversification in any of the species. The Arabidopsis gene (At4g00340.1) has not been characterized, but its ortholog in strawberry (*M2F10*) is upregulated in response to infection by the fungus *Colletotrichum acutatum* [87]. Given that many other G-LecRKs respond to fungal pathogens, it is difficult to hypothesize why this particular clade has not undergone the type of expansion seen in nearly every other clade.

The sister clade to B2a (B2b, with 95% support) is subdivided into 5 clades that are unresolved relative to each other. One of these (B2b1) also shows relatively little expansion: it consists of two sister-clades, each of which has genes from all three species. This indicates a duplication before the diversification of the eudicots, but the only further expansion is a single columbine duplication. The two Arabidopsis genes in clade B2b1 respond to abiotic stress: At2g41890.1

is downregulated in response to gravity, and At1g34300.1, which responds to water loss and decreased dry weight [72, 74]. This suggests a role in abiotic stress response for the tomato and columbine genes as well.

Clade B2b contains two larger subclades that show greater diversification. Clade B2b2 (100% support) has one Arabidopsis member, At5g60900.1, which is implicated in both biotic (upregulated by *Fusarium* and LPS) and abiotic (downregulated by cold) responses [72, 78, 88]. Although resolution within this clade is poor, the nested position of the Arabidopsis gene suggests that the low number of genes from this species may be the result of loss. However, further resolution could show instead independent diversification in tomato and columbine. Clade B2b3 (99% support) is subdivided into two sister clades, one of which consists entirely of tomato and columbine genes, strongly suggesting loss in Arabidopsis. The other subclade consists of two sister-clades each of which has one gene from each species, another example of a single duplication followed by no additional diversification in any species. One of the Arabidopsis genes, *SD2-5* (At4g32300.1), is not characterized, but the rice homolog, *Pi-d2*, confers resistance to the fungal pathogen *Magnaporthe grisea* [89]. The other Arabidopsis gene, At5g35370.1, is downregulated in response to nematode infection and is important in salt and drought tolerance, thereby showing roles in both biotic and abiotic responses [90, 91].

Kinase domain analysis of Arabidopsis G-LecRKs

The presence of the 11 known kinase subdomains and sites essential for the catalytic activity [51] of the G-LecRKs was investigated by aligning the amino acid sequences of the kinase domains to the LecRK-VI.2 kinase domain [13]. The alignment revealed overall conservation of the ATP binding (consensus motif G-x-G-x-x-G-x-V) and the catalytic sites (consensus motif H-R-D-L-K-x-x-N), with a few substitutions in the other kinase subdomains (Additional file 5) [92]. Two of the G-LecRK sequences, At1g67520 and At5g24080, revealed incomplete kinase domains, lacking four (VIII – XI) and seven (I–VII) of the 11 kinase subdomains, respectively. Additionally, At2g41890 has several mutations and deletions notably, in the ATP binding site, in subdomain I, as well as a single amino acid change in its catalytic loop in subdomain VI (Additional file 5) [92]. The essential arginine and aspartic acid residues are substituted for glycine and asparagine, respectively. Therefore, At1g67520, At5g24080, and At2g41890 are likely inactive kinases.

Nomenclature for the Arabidopsis G-LecRKs

The L-LecRKs were previously classified and a nomenclature was established based on the amino acid and nucleotide sequences of the 45 members of the family [34].

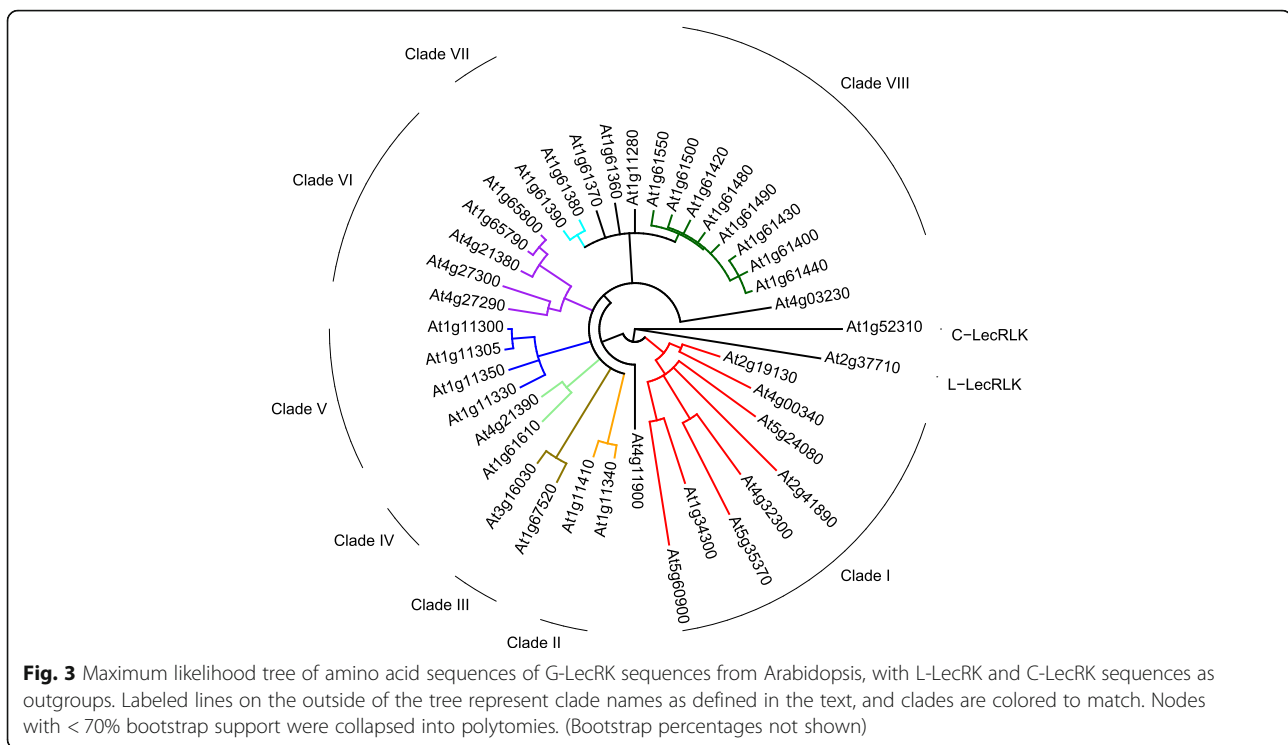
In that system, clades, defined as groups of genes with at least 50% identity among homologs at both the nucleotide and amino acid level, were designated by Roman numerals [34]. Following a similar approach, but basing clade membership on the results of our phylogenetic analysis rather than percent sequence identity, we classified the 38 members of the Arabidopsis G-LecRK group into eight clades (I–VIII) (Fig. 3, Table 3, Additional file 6). Within each clade, genes are numbered progressively, across chromosomes and clades as in Bouwmeester et al. [34]; therefore, gene #1 in each clade would be the gene on the lowest numbered chromosome with the lowest numerical location (for clade I, At1G34300, designated G-LecRK-I.1) and the final gene would be the one on the highest numbered chromosome with the highest numerical location (for clade I, At5G60900, designated G-LecRK-I.8). The two largest identified clades (clades I and VIII) contain eight members, followed by clades with five and four members (VI and V, respectively) and four clades with two members each (Fig. 3, Table 3). Five genes were not placed in any clade, behaving as singletons. Chromosomal location was not predictive of clade membership, with genes from chromosome 1 being found in all eight clades; in addition, three of the singleton genes are located on chromosome 1 where the vast majority of G-LecRKs are localized.

Chromosomal location and prediction of Arabidopsis

G-LecRK subcellular localization

The 38 Arabidopsis G-LecRKs were mapped onto the five chromosomes using the chromosomal map tool from TAIR. A single G-LecK locus, At1g11305, was not present on TAIR since it was created when At1g11300 was discovered as a misannotated gene and was split into At1g11300 and At1g11305 [70]. Therefore, the chromosomal location of At1g11305 was based on the location of At1g11300. Unlike Arabidopsis L-LecRKs, most of which are localized on chromosomes 5 and 3 [28, 30], the vast majority of the Arabidopsis G-LecRKs are localized on chromosome 1 (24 members), followed by chromosome 4 (eight members), chromosome 5 (three members), chromosome 2 (two members) and chromosome 3 (one member) (Additional file 7).

Arabidopsis G-LecRKs localization was predicted using SUBA3 [58, 59]. This tool predicted all Arabidopsis proteins to be localized at the plasma membrane, consistent with the existence of a TM domain. SUBA predictions were further investigated with TargetP 1.1 [60]. This tool predicts protein localization by analyzing cleavage site predictions and, therefore, predicts localization to the chloroplast, mitochondria or secretory pathways. Most of the Arabidopsis G-LecRKs were predicted to have a secretion signal peptide (Additional file 8). Two proteins, At1g61390 and At1g61400,



were predicted to localize at the mitochondrial membrane and localization was not predicted by Target P 1.1 for one, At1g11280.

To validate localization predictions, the subCELLular LOcalization tool CELLO [61] was used. CELLO predictions mostly confirmed the predictions obtained by SUBA, but additionally revealed possible specific subcellular localization of two G-LecRKs, At4g27290 and At5g60900 (Additional file 8). These encode proteins without TMs, based on a domain search performed using Interpro, although both proteins were predicted to localize at the plasma membrane by SUBA. Interestingly, CELLO prediction added the possibility that these proteins could also localize to the nucleus and cytoplasm (Additional file 8).

Kinase domain analysis of the tomato G-LecRKs

Like Arabidopsis, the presence of sites essential for catalytic activities of the 11 kinase subdomains [92] was investigated for the tomato G-LecRKs. The alignment of the tomato G-LecRKs kinase domains revealed overall conservation of the ATP-binding and catalytic sites, with a few substitutions in the other kinase subdomains (Additional file 9). The search also revealed ten genes with incomplete kinase domains with various amino acid modifications and indels in the subdomains (Table 4, Additional file 9). Additionally, Solyc07g063810 shows conservation of subdomains VI to XI, which includes the catalytic site, but displays several amino acid modifications

in subdomains I to V, including the ATP binding site, suggesting it is likely an inactive kinase. Solyc03g063650 has a substitution of the aspartic acid to asparagine the kinase catalytic site, in subdomain VI, and lacks essential amino acids of subdomains I to IV suggesting it is also likely an inactive kinase (Additional file 9).

Nomenclature for the tomato G-LecRKs

Following a similar nomenclature as for the Arabidopsis G-LecRKs, the clades containing the 73 tomato G-LecRK members were used as the basis for naming the genes. This methodology resulted in the grouping of tomato G-LecRKs into 13 clades, within which genes are numbered progressively across chromosomes and clades (Fig. 4, Table 5, Additional file 10). Clades range in size from the largest, clade III, with 21 members, to four clades with two members each. Three genes do not fall in any of these clades and are designated singletons, two of which are on chromosome 2 and one on chromosome 7. As with Arabidopsis, chromosomal location is not correlated with clade membership.

Chromosomal location and prediction of tomato G-LecRKs subcellular localization

The 73 G-LecRK loci were mapped onto the 12 tomato chromosomes. As for Arabidopsis, locations of the genes that were misannotated and split into 2, were based on the location of the original locus. Members of tomato G-LecRKs are distributed throughout the 12 tomato chromosomes, with over

Table 3 Classification and proposed nomenclature of the Arabidopsis G-LecRKs

Clade name	Gene name	Locus
G-LecRK-I	<i>AtG-LecRK-I.1</i>	AT1G34300
	<i>AtG-LecRK-I.2</i>	AT2G19130
	<i>AtG-LecRK-I.3</i>	AT2G41890
	<i>AtG-LecRK-I.4</i>	AT4G00340
	<i>AtG-LecRK-I.5</i>	AT4G32300
	<i>AtG-LecRK-I.6</i>	AT5G24080
	<i>AtG-LecRK-I.7</i>	AT5G35370
	<i>AtG-LecRK-I.8</i>	AT5G60900
G-LecRK-II	<i>AtG-LecRK-II.1</i>	AT1G11340
	<i>AtG-LecRK-II.2</i>	AT1G11410
G-LecRK-III	<i>AtG-LecRK-III.1</i>	AT1G67520
	<i>AtG-LecRK-III.2</i>	AT3G16030
G-LecRK-IV	<i>AtG-LecRK-IV.2</i>	AT1G61610
	<i>AtG-LecRK-IV.2</i>	AT4G21390
G-LecRK-V	<i>AtG-LecRK-V.1</i>	AT1G11300
	<i>AtG-LecRK-V.2</i>	AT1G11305
	<i>AtG-LecRK-V.3</i>	AT1G11330
	<i>AtG-LecRK-V.4</i>	AT1G11350
G-LecRK-VI	<i>AtG-LecRK-VI.1</i>	AT1G65790
	<i>AtG-LecRK-VI.2</i>	AT1G65800
	<i>AtG-LecRK-VI.3</i>	AT4G21380
	<i>AtG-LecRK-VI.4</i>	AT4G27290
	<i>AtG-LecRK-VI.5</i>	AT4G27300
G-LecRK-VII	<i>AtG-LecRK-VII.1</i>	AT1G61380
	<i>AtG-LecRK-VII.2</i>	AT1G61390
G-LecRK-VIII	<i>AtG-LecRK-VIII.1</i>	AT1G61400
	<i>AtG-LecRK-VIII.2</i>	AT1G61420
	<i>AtG-LecRK-VIII.3</i>	AT1G61430
	<i>AtG-LecRK-VIII.4</i>	AT1G61440
	<i>AtG-LecRK-VIII.5</i>	AT1G61480
	<i>AtG-LecRK-VIII.6</i>	AT1G61490
	<i>AtG-LecRK-VIII.7</i>	AT1G61500
	<i>AtG-LecRK-VIII.8</i>	AT1G61550
Singletons	<i>AtG-LecRK-S.1</i>	AT1G11280
	<i>AtG-LecRK-S.2</i>	AT1G61360
	<i>AtG-LecRK-S.3</i>	AT1G61370
	<i>AtG-LecRK-S.4</i>	AT4G03230
	<i>AtG-LecRK-S.5</i>	AT4G11900

half (54%) localized on chromosome 7, 4, 2, and 3 encompassing 18, 14, 11, and 11 members, respectively (Additional file 11) [30]. In contrast, tomato L-LecRKs are mostly localized on chromosome 9 and 10, with members located on 8 of the 12 chromosomes.

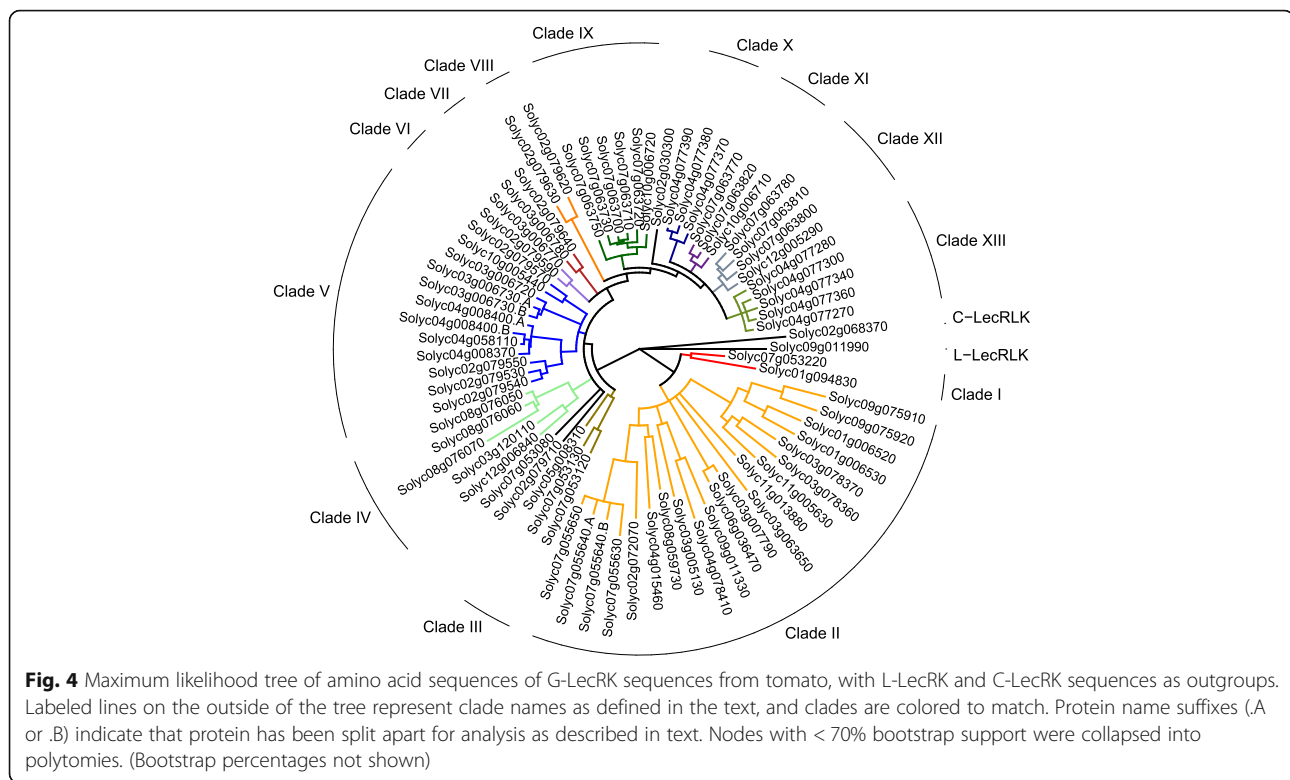
Table 4 Tomato G-LecRKs with incomplete kinase subdomains that lack some of the 11 subdomains

Tomato G-LecRK	Present kinase subdomains
Solyc04g008400.B	I and II
Solyc03g006780	I - V
Solyc04g008370	I - V
Solyc04g077300	I - V
Solyc08076070	I - V, VI
Solyc07g055630	I, II, VI - X
Solyc04g077380	I - V, VI, XI
Solyc07g055640.A	I - X
Solyc02g079710	I - X
Solyc07g063750	I - X

Tomato G-LecRK protein sequences were used to predict subcellular localization with TargetP 1.1 [60]. The localization of tomato C-LecRKs and L-LecRKs have not been investigated. However L-LecRKs possess TM domains and are predicted to localize mostly at the plasma membrane with a few members predicted to localize to mitochondria or chloroplast [28]. TargetP predicted that the great majority of tomato G-LecRKs have secretion pathway signals and are localized to the plasma membrane (Additional file 12). A single protein (Solyc02g079630) was predicted to have chloroplast localization. Four proteins (Solyc03g006730.B, Solyc07g055640.A, Solyc07g063810 and Solyc11g005630) were predicted to localize to mitochondria. Eight proteins (Solyc07g055640.B, Solyc04g077380, Solyc08g076060, Solyc07g055650, Solyc07g055630, Solyc04g008400.B, Solyc02g030300 and Solyc08g076070), for which a signal peptide could not be predicted using this tool, were not localized to a specific subcellular compartment. Additionally, the subcellular localization tool, CELLO, was used to investigate the localization of the tomato G-LecRKs and showed an overlap of prediction of plasma membrane localization and presence of TargetP secretion pathway signal. Interestingly, this tool was able to predict subcellular localization of proteins for which TargetP could not predict localization, and was also able to predict membrane localization for proteins that did not have a predicted TM domain, suggesting a different membrane-associated signal for these proteins. CELLO predictions also suggested multiple localizations for a few tomato G-LecRKs (such as Solyc01g006530 and Solyc07g055630) and contradicted a few predictions by TargetP (such as Solyc02g079630 and Solyc03g006730.B; Additional file 12).

Comparison of conserved motifs in ectodomains of tomato and Arabidopsis G-LecRKs

The predicted cytoplasmic-localized regions of G-LecRKs consist of the extremely conserved kinase domains. To



investigate the presence of conserved motifs in the ectodomain of the Arabidopsis and tomato G-LecRKs, the amino acid sequences of the ectodomains were submitted to MEME [57]. Despite the high variability of the ectodomains, five motifs present in at least 30 of the 38 Arabidopsis sequences and in at least 45 of the 73 tomato sequences were identified (Fig. 5).

The highest conserved ectodomain motif (Fig. 5a) was present in all Arabidopsis and tomato G-LecRKs and it was previously shown to be present in 96% of the rice G-LecRKs [28]. One of the motifs, a cysteine-rich region within the PAN domain (Fig. 5b), is present in 34 and 66 of the Arabidopsis and tomato G-LecRKs, respectively. Interestingly, a third motif (Fig. 5c) is also conserved in 76 out of 100 rice G-LecRKs, and was previously identified in 27 Arabidopsis G-LecRKs [28]. The two remaining ectodomain motifs are novel and have not been reported previously (Fig. 5d, e). The conservation of the motifs in the ectodomain of both Arabidopsis and tomato G-LecRKs is remarkable considering that these extracellular domains harbor the lectin domain known to have low conservation among members of this family from a single plant species [28].

Discussion

Two previous studies have reported Arabidopsis to have 32 [28, 93] G-LecRK members, a number lower than the 38 members identified in this study. One of the possible reasons for this discrepancy from Shiu

and Bleecker [93], is the current improved annotation of the Arabidopsis genome. Their study also used the presence of a TM domain as a criterion for their analysis, but this was not used in our study. As for the lower number identified by Vaid et al. [28], this could be due to the fact that their analysis relied on sequence similarity to a single gene sequence, At1g61610, while in our analysis, we used a number of candidate sequences from our initial search results to fish out additional candidates. Nevertheless, their overall criterion for candidates was the same as the one used here, which is the presence of both lectin and kinase domains.

Our search retrieved all genes identified by Vaid et al. [28] and an additional 16 sequences that have a G-type lectin domain. Their gene list included a single gene, At1g61460, that is not present on our list and which does not encode a G-type lectin domain, so was not recovered in our BLASTp searches. Of the 16 new sequences with a G-type lectin domain, nine do not encode a kinase domain and would not have been retrieved by Vaid et al. [28]. These were also excluded from our analysis. Taken together, our results added seven proteins to the previous list of Arabidopsis G-LecRKs. Of these seven additional genes, At1g67520 and At5g24080 encode proteins with atypical kinase domains and lacking several subdomains, while At2g41890 lacked essential amino acids at the ATP binding site and the catalytic

Table 5 Classification and proposed nomenclature of the tomato G-LecRKs

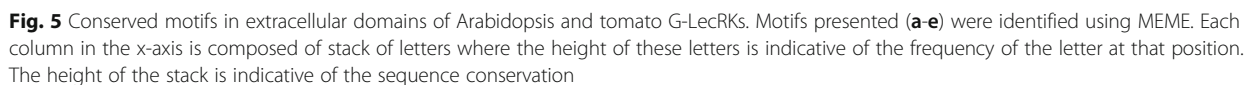
Clade name	Gene name	Locus
G-LecRK-I	<i>SIG-LecRK-I.1</i>	Solyc01g094830
	<i>SIG-LecRK-I.2</i>	Solyc07g053220
G-LecRK-II	<i>SIG-LecRK-II.1</i>	Solyc01g006520
	<i>SIG-LecRK-II.2</i>	Solyc01g006530
	<i>SIG-LecRK-II.3</i>	Solyc02g072070
	<i>SIG-LecRK-II.4</i>	Solyc03g005130
	<i>SIG-LecRK-II.5</i>	Solyc03g007790
	<i>SIG-LecRK-II.6</i>	Solyc03g063650
	<i>SIG-LecRK-II.7</i>	Solyc03g078370
	<i>SIG-LecRK-II.8</i>	Solyc03g078360
	<i>SIG-LecRK-II.9</i>	Solyc04g015460
	<i>SIG-LecRK-II.10</i>	Solyc04g078410
	<i>SIG-LecRK-II.11</i>	Solyc06g036470
	<i>SIG-LecRK-II.12</i>	Solyc07g055630
	<i>SIG-LecRK-II.13</i>	Solyc07g055640.A
	<i>SIG-LecRK-II.14</i>	Solyc07g055640.B
	<i>SIG-LecRK-II.15</i>	Solyc07g055650
	<i>SIG-LecRK-II.16</i>	Solyc08g059730
	<i>SIG-LecRK-II.17</i>	Solyc09g011330
	<i>SIG-LecRK-II.18</i>	Solyc09g075910
	<i>SIG-LecRK-II.19</i>	Solyc09g075920
	<i>SIG-LecRK-II.20</i>	Solyc11g005630
	<i>SIG-LecRK-II.21</i>	Solyc11g013880
G-LecRK-III	<i>SIG-LecRK-III.1</i>	Solyc05g008310
	<i>SIG-LecRK-III.2</i>	Solyc07g053120
	<i>SIG-LecRK-III.3</i>	Solyc07g053130
G-LecRK-IV	<i>SIG-LecRK-IV.1</i>	Solyc03g120110
	<i>SIG-LecRK-IV.2</i>	Solyc08g076050
	<i>SIG-LecRK-IV.3</i>	Solyc08g076060
	<i>SIG-LecRK-IV.4</i>	Solyc08g076070
	<i>SIG-LecRK-IV.5</i>	Solyc12g006840
G-LecRK-V	<i>SIG-LecRK-V.1</i>	Solyc02g079530
	<i>SIG-LecRK-V.2</i>	Solyc02g079540
	<i>SIG-LecRK-V.3</i>	Solyc02g079550
	<i>SIG-LecRK-V.4</i>	Solyc02g079570
	<i>SIG-LecRK-V.5</i>	Solyc03g006720
	<i>SIG-LecRK-V.6</i>	Solyc03g006730.A
	<i>SIG-LecRK-V.7</i>	Solyc03g006730.B
	<i>SIG-LecRK-V.8</i>	Solyc04g008370
	<i>SIG-LecRK-V.9</i>	Solyc04g008400.A
	<i>SIG-LecRK-V.10</i>	Solyc04g008400.B
	<i>SIG-LecRK-V.11</i>	Solyc04g058110
	<i>SIG-LecRK-V.12</i>	Solyc10g005440

Table 5 Classification and proposed nomenclature of the tomato G-LecRKs (Continued)

Clade name	Gene name	Locus
G-LecRK-VI	<i>SIG-LecRK-VI.1</i>	Solyc02g079590
	<i>SIG-LecRK-VI.2</i>	Solyc03g006770
G-LecRK-VII	<i>SIG-LecRK-VII.1</i>	Solyc02g079640
	<i>SIG-LecRK-VII.2</i>	Solyc03g006780
G-LecRK-VIII	<i>SIG-LecRK-VIII.1</i>	Solyc02g079620
	<i>SIG-LecRK-VIII.2</i>	Solyc02g079630
G-LecRK-IX	<i>SIG-LecRK-IX.1</i>	Solyc07g063700
	<i>SIG-LecRK-IX.2</i>	Solyc07g063710
	<i>SIG-LecRK-IX.3</i>	Solyc07g063720
	<i>SIG-LecRK-IX.4</i>	Solyc07g063730
	<i>SIG-LecRK-IX.5</i>	Solyc07g063750
	<i>SIG-LecRK-IX.6</i>	Solyc10g006720
G-LecRK-X	<i>SIG-LecRK-X.1</i>	Solyc04g077370
	<i>SIG-LecRK-X.2</i>	Solyc04g077380
	<i>SIG-LecRK-X.3</i>	Solyc04g077390
G-LecRK-XI	<i>SIG-LecRK-XI.1</i>	Solyc07g063770
	<i>SIG-LecRK-XI.2</i>	Solyc07g063820
	<i>SIG-LecRK-XI.3</i>	Solyc10g006710
G-LecRK-XII	<i>SIG-LecRK-XII.1</i>	Solyc07g063780
	<i>SIG-LecRK-XII.2</i>	Solyc07g063800
	<i>SIG-LecRK-XII.3</i>	Solyc07g063810
	<i>SIG-LecRK-XII.4</i>	Solyc12g005290
G-LecRK-XIII	<i>SIG-LecRK-XIII.1</i>	Solyc04g077270
	<i>SIG-LecRK-XIII.2</i>	Solyc04g077280
	<i>SIG-LecRK-XIII.3</i>	Solyc04g077300
	<i>SIG-LecRK-XIII.4</i>	Solyc04g077340
	<i>SIG-LecRK-XIII.5</i>	Solyc04g077360
Singletons	<i>SIG-LecRK-S.1</i>	Solyc02g030300
	<i>SIG-LecRK-S.2</i>	Solyc02g079710
	<i>SIG-LecRK-S.3</i>	Solyc07g053080

loop, suggesting they are defective kinases [34]. While kinase activity could be crucial for the function of these proteins, lacking kinase activity may not abolish function, as function for kinase inactive receptor-like kinases has been previously reported [94, 95].

The same search methodology used for identifying members of this family in Arabidopsis showed success with identification of the members in tomato. These investigations of the tomato genome allowed the identification of 73 genes encoding proteins with both a kinase domain and a G-lectin domain and revealed expansion of the number of members of this family in tomato relative to Arabidopsis. As in Arabidopsis, several of the 73 tomato G-LecRKs have mutations and deletions in their



Phylogenetic analysis of G-LecRK sequences consistently shows the genes falling into two clades, whether one, three, or four species is included. One of these clades includes the L-LecRK sequences intended as out-group along with the C-LecRKs. This configuration would indicate that the L-type genes should be merged into the G-type, because sequence analysis shows they are not distinct. However, *Arabidopsis* alone has 45 L-LecRKs and this analysis included only one, therefore

There is no pattern discernable in the expression and function data currently available regarding G-LecRKs.

Whereas all recorded functions and expression data are consistent with roles in biotic and abiotic stress responses or developmental processes, which involve similar pathways, no clade seems specialized for specific functions within these categories. It is not possible to say “there is a clade of genes that responds to trehalose, and it has expanded in species X, therefore trehalose response is important in species X.” Rather, we find closely related genes that appear to have taken on different functions related to stress response and development. Clade A2b2, which includes the largest cluster of Arabidopsis genes, includes members with putative functions ranging from gravitropism to pollen and root hair development to resistance to bacterial infection [71, 75, 76, 84, 85, 96]. Clades with fewer Arabidopsis sequences, such as B2b1 with two, still show variation, in this case one gene that responds to gravity and one that responds to water loss [72, 74]. Two points are important to keep in mind when evaluating these expression and function data, however. First, data are only available for genes from Arabidopsis; we currently do not have data for tomato or columbine genes. Second, the current analyses are based on published reports, but it is highly likely that no study tested all possible biotic and abiotic factors. Therefore the fact that At1g11340.1 is suppressed by mevalonic acid does not mean it might not be up- or downregulated in response to other factors [73].

Analysis of the clades in the phylogenetic tree shows that there has been species-specific gene expansion in different clades across the tree, and examination of chromosomal locations of the genes suggests combinations of tandem and possible whole genome duplications. Notably, clade B has only 7 Arabidopsis genes, and no subclade has more than two. In contrast, clade B has 23 tomato genes, with one subclade that has a single Arabidopsis gene (B2b2) having 7 tomato genes. This same subclade has 5 columbine genes. Clade B1 has one gene each from Arabidopsis and tomato, but 9 from columbine. Clade A has a greater number of genes from Arabidopsis, but again we see species-specific expansion in different subclades in all three species. Clade A2b2 has 3 tomato genes, which form a sister-group to 13 Arabidopsis genes; this is the largest Arabidopsis gene group in the tree. The tomato genes are all located on chromosome 3, and the Arabidopsis genes are all located on chromosome 1, suggesting a single ancestral gene that underwent tandem duplications in each species independently, and to a greater extent in Arabidopsis. Clade A1 consists of three subclades, two of which have genes only from Arabidopsis and tomato, and one of which has genes from all three. This topology implies three ancestral genes, with loss of columbine genes from the first two clades. These clades

(A1a, A1b) further show evidence of additional duplications within tomato and Arabidopsis, although in this case these appear potentially to be a combination of tandem and whole genome duplications based on chromosomal locations. The third clade, A1c, again shows dramatic expansion in columbine and especially tomato, but not Arabidopsis. Because there is no clear pattern of differential function or expression among clades, it is possible that the differential diversification of clades in different species is essentially stochastic. Duplication appears to be extremely common, and the extent in each clade in each species may not be specifically under selection. In fact, the independent expansion and relatively small number of clades with genes from both tomato and Arabidopsis suggests that the base number of core eudicot G-LecRKs is low, and that the differences in numbers between the two species is more a result of expansion than loss. This is also consistent with the role of these genes in stress responses.

While clustering of G-LecRKs members on chromosomes, such as Arabidopsis chromosome 1 and tomato chromosome 7, suggests duplications, our analysis indicates that chromosomal location is not predictive of a clade membership for either plant species. This suggests that G-LecRKs are rapidly evolving and diversifying consistent with their known functional roles in biotic and abiotic stress responses and development.

Clade membership was also not indicative of the presence of specific ectodomain configurations, specifically presence (or absence) of three domains: SLG, EGF and PAN. The importance of each of these domains, as well as their contributions to G-LecRK activity, have not been investigated to date. Nevertheless, it is to be expected that relevant regions at the ectodomain, outside of the region that confers substrate-binding specificity, would be conserved among different members of the same family. Consistent with this hypothesis, a motif search among members from Arabidopsis and tomato revealed the presence of a single motif (Fig. 5a) in all members of G-LecRKs from both plant species. This motif was also identified in a previous investigation in both Arabidopsis and rice (96% of rice G-LecRKs) [28]. Interestingly, the second motif identified in our search (Fig. 5b), present in 34 Arabidopsis G-LecRKs and 66 tomato G-LecRKs was also identified in 76% of the rice G-LecRKs by the same authors. The observation that the lectin domain is the domain with low conservation in G-LecRKs and the presence of conserved motifs in the ectodomain shows that despite the lack of conservation of the lectin domain, a specific motif is conserved and might constitute essential site(s) for protein activity. The two newly identified motifs, observed in 30 of the 38 Arabidopsis and

45 of the 73 tomato lectin domain of G-LecRKs, are less common. Their presence in additional plant species and how widespread these new motifs are remains to be investigated.

Conclusions

We present here the results of an analysis of G-LecRK gene lineage evolution in Arabidopsis, a member of the mustard family (Brassicaceae, in the rosid clade) and tomato, a member of the nightshade family (Solanaceae, in the asterid clade). Given that these are both members of the derived angiosperm clade, core eudicots, we also included an evaluation of G-LecRKs in columbine (*Aquilegia coerulea*), a member of the basal eudicots that diverged before the rosid-asterid split in the core eudicots to polarize the tree and allow analysis of duplication, expansion, and loss of G-LecRK genes and gene clades.

The present investigation added to the number of currently known Arabidopsis G-LecRKs and characterized for the first time the tomato G-LecRKs. We proposed a nomenclature for both Arabidopsis and tomato G-LecRKs and identified possible essential sites for G-LecRK activity. Additionally, prediction of protein localization by different tools enriched the initial prediction of G-LecRKs plasma membrane localization and raised the possibility for specificity of modes of actions of a number of G-LecRKs depending on their specific subcellular localization patterns. Given their putative roles in plant defense, and the importance of tomato as a crop, an understanding of the structure and evolution of these proteins in tomato may shed light on defense strategies that can be leveraged to produce hardier plants and yield.

Additional files

Additional file 1: Members of the columbine G-lectins and their domains. (XLSX 57 kb)

Additional file 2: Members of the rice G-lectins and their domains. (XLS 33 kb)

Additional file 3: Maximum likelihood tree of amino acid sequences from G-LecRK, L-LecRK intended outgroups, and C-LecRK outgroups from tomato, Arabidopsis, columbine, and rice. Bootstrap support from 1000 replicates is shown above nodes. Brackets on the right indicate intended outgroup clades. (PDF 20 kb)

Additional file 4: Maximum likelihood tree of amino acid sequences from G-LecRK, L-LecRK intended outgroups, and C-LecRK outgroups from tomato, Arabidopsis, and columbine. Bootstrap support from 1000 replicates is shown above nodes. Brackets on the right indicate the major clades as defined in the text along with the two intended outgroup clades. (PDF 19 kb)

Additional file 5: Alignment of predicted amino acid sequences of Arabidopsis G-LecRKs kinase domains with the L-LecRK-VI.2 using ClustalW. Lines on top of the alignment show subdomains I and II, ATP binding site (GxGxxGxV) and subdomain VI, the serine/threonine kinase active site (HRDLKxxN). (PDF 3979 kb)

Additional file 6: Maximum likelihood tree of amino acid sequences from G-LecRK, L-LecRK outgroups, and C-LecRK outgroups from

Arabidopsis. Bootstrap support from 1000 replicates is shown above nodes. (PDF 11 kb)

Additional file 7: Genetic map of the Arabidopsis G-LecRKs. Arrangement of G-LecRKs on the five Arabidopsis chromosomes. Figure was prepared using Chromosome Map Tool in TAIR. Locus At1g11305 was added manually. (TIFF 337 kb)

Additional file 8: Subcellular localization of the Arabidopsis G-LecRKs. Localization was predicted using SUBA, TargetP and CELLO software programs/tools. (PDF 55 kb)

Additional file 9: Alignment of predicted amino acid sequences of tomato G-LecRKs kinase domains with Solyc03g006720 using ClustalW. Lines on top of alignment show subdomains I and II, ATP binding site (GxGxxGxV) and subdomain VI, the serine/threonine kinase active site (HRDLKxxN). (PDF 7110 kb)

Additional file 10: Maximum likelihood tree of amino acid sequences from G-LecRK, L-LecRK outgroups, and C-LecRK outgroups from tomato. Bootstrap support from 1000 replicates is shown above nodes. (PDF 7 kb)

Additional file 11: Genetic map of the tomato G-LecRKs. Arrangement of G-LecRKs on the 12 tomato chromosomes. Figure was prepared manually using map viewer in NCBI. (TIFF 309 kb)

Additional file 12: Subcellular localization of the tomato G-LecRKs. Localization was predicted using TargetP and CELLO software programs/tools. (PDF 56 kb)

Abbreviations

C-type: Calcium-dependent; EGF: Epidermal growth factor; G-type: *Galanthus nivalis* agglutinin type; LecRKs: Lectin receptor kinases; LPS: Lipopolysaccharide; LRR: Leucine-rich repeat; L-type: Legume-like; PAN: Plasminogen-apple-nematode; PRR: Pattern-recognition receptors; *Pst*: *Pseudomonas syringae* pv. *tomato*; RK: Receptor kinase; SLG: S-locus glycoprotein; TM: Transmembrane

Acknowledgements

None

Funding

MAT was funded by the Coordination for the Improvement of Higher Education of Brazil. OGC was funded by a grant from National Science Foundation Research Experiences for Undergraduates (1461297).

Availability of data and materials

Datasets derived from public resources and made available with the article.

Authors' contributions

MAT, AR, AL and IK designed the analyses. MAT, AR, JH and OGC performed the analyses. MAT, AR, AL and IK wrote the manuscript. All authors read and approved the final version of the manuscript.

Ethics approval and consent to participate

This study has not directly involved humans, animals or plants.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Nematology, University of California, Riverside, California, USA. ²Department of Botany and Plant Sciences, University of California, Riverside, California, USA. ³Chaffey College, Rancho Cucamonga, California, USA. ⁴Institute for Integrative Genome Biology, University of California, Riverside, California, USA.

Received: 9 June 2017 Accepted: 16 March 2018

Published online: 06 April 2018

References

- Lehti-Shiu MD, Shiu S-H. Diversity, classification and function of the plant protein kinase superfamily. *Philos Trans R Soc Lond Ser B Biol Sci*. 2012;367:2619–39.
- Gilardini PA, Hettenhausen C, Baldwin IT, Bonaventure G. *Nicotiana attenuata* LECTIN RECEPTOR KINASE1 suppresses the insect-mediated inhibition of induced defense responses during *Manduca sexta* herbivory. *Plant Cell*. 2011;23:3512–32.
- Bouwmeester K, de Sain M, Weide R, Gouget A, Klammer S, Canut H, et al. The lectin receptor kinase LecRK-I.9 is a novel *Phytophthora* resistance component and a potential host target for a RXLR effector. *PLoS Pathog*. 2011;7:e1001327.
- Arnaud D, Desclos-Theveniau M, Zimmerli L. Disease resistance to *Pectobacterium carotovorum* is negatively modulated by the Arabidopsis lectin receptor kinase LecRK-V.5. *Plant Signal Behav*. 2012;7:1070–2.
- Desclos-Theveniau M, Arnaud D, Huang T-Y, Lin G-J, Chen W-Y, Lin Y-C, et al. The Arabidopsis lectin receptor kinase LecRK-V.5 represses stomatal immunity induced by *Pseudomonas syringae* pv. *tomato* DC3000. *PLoS Pathog*. 2012;8:e1002513.
- Singh P, Kuo Y-C, Mishra S, Tsai C-H, Chien C-C, Chen C-W, et al. The lectin receptor kinase-VI.2 is required for priming and positively regulates Arabidopsis pattern-triggered immunity. *Plant Cell*. 2012;24:1256–70.
- Armijo G, Salinas P, Monteoliva MI, Seguel A, García C, Villarroel-Candia E, et al. A salicylic acid-induced lectin-like protein plays a positive role in the effector-triggered immunity response of *Arabidopsis thaliana* to *Pseudomonas syringae* Avr-Rpm1. *Mol Plant-Microbe Interact*. 2013;26:1395–406.
- Cheng X, Wu Y, Guo J, Du B, Chen R, Zhu L, et al. A rice lectin receptor-like kinase that is involved in innate immune responses also contributes to seed germination. *Plant J*. 2013;76:687–98.
- Singh P, Zimmerli L. Lectin receptor kinases in plant innate immunity. *Front Plant Sci*. 2013;4:124.
- Lannoo N, Van Damme EJM. Lectin domains at the frontiers of plant defense. *Front Plant Sci*. 2014;5:397.
- Macedo MLR, Oliveira CFR, Oliveira CT. Insecticidal activity of plant lectins and potential application in crop protection. *Molecules*. 2015;20:2014–33.
- Bigard J, Colcombet J, Hirt H. Signaling mechanisms in pattern-triggered immunity (PTI). *Mol Plant*. 2015;8:521–39.
- Singh P, Chien C-C, Mishra S, Tsai C-H, Zimmerli L. The Arabidopsis LECTIN RECEPTOR KINASE-VI.2 is a functional protein kinase and is dispensable for basal resistance to *Botrytis cinerea*. *Plant Signal Behav*. 2013;8:e22611.
- Zipfel C. Plant pattern-recognition receptors. *Trends Immunol*. 2014;35:345–51.
- Malinovskiy FG, Fangel JU, Willats WGT. The role of the cell wall in plant immunity. *Front Plant Sci*. 2014;5:178.
- Böhm H, Albert I, Fan L, Reinhard A, Nürnberger T. Immune receptor complexes at the plant cell surface. *Curr Opin Plant Biol*. 2014;20:47–54.
- Felix G, Duran JD, Volko S, Boller T. Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant J*. 1999;18:265–76.
- Trdál L, Fernandez O, Boutrot F, Héloir M-C, Kelloniemi J, Daire X, et al. The grapevine flagellin receptor VvFLS2 differentially recognizes flagellin-derived epitopes from the endophytic growth-promoting bacterium *Burkholderia phytofirmans* and plant pathogenic bacteria. *New Phytol*. 2014;201:1371–84.
- Shi Q, Febres VJ, Jones JB, Moore GA. A survey of FLS2 genes from multiple citrus species identifies candidates for enhancing disease resistance to *Xanthomonas citri* ssp. *citri*. *Hortic Res*. 2016;3:16022.
- Robatzek S, Bittel P, Chinchilla D, Köchner P, Felix G, Shiu S-H, et al. Molecular identification and characterization of the tomato flagellin receptor LeFLS2, an orthologue of Arabidopsis FLS2 exhibiting characteristically different perception specificities. *Plant Mol Biol*. 2007;64:539–47.
- Takai R, Isogai A, Takayama S, Che F-S. Analysis of flagellin perception mediated by flg22 receptor OsFLS2 in rice. *Mol Plant-Microbe Interact*. 2008;21:1635–42.
- Hind SR, Strickler SR, Boyle PC, Dunham DM, Bao Z, O'Doherty IM, et al. Tomato receptor FLAGELLIN-SENSING 3 binds flgII-28 and activates the plant immune system. *Nat Plants*. 2016;2:16128.
- Heese A, Hann DR, Gimenez-Ibanez S, Jones AME, He K, Li J, et al. The receptor-like kinase SERK3/BAK1 is a central regulator of innate immunity in plants. *Proc Natl Acad Sci U S A*. 2007;104:12217–22.
- Ron M, Avni A. The receptor for the fungal elicitor ethylene-inducing xylanase is a member of a resistance-like gene family in tomato. *Plant Cell*. 2004;16:1604–15.
- Cao Y, Liang Y, Tanaka K, Nguyen CT, Jedrzejczak RP, Joachimiak A, et al. The kinase LYK5 is a major chitin receptor in Arabidopsis and forms a chitin-induced complex with related kinase CERK1. *Elife*. 2014;3 Available from: <https://doi.org/10.7554/eLife.03766>
- Shinya T, Motoyama N, Ikeda A, Wada M, Kamiya K, Hayafune M, et al. Functional characterization of CEBiP and CERK1 homologs in Arabidopsis and rice reveals the presence of different chitin receptor systems in plants. *Plant Cell Physiol*. 2012;53:1696–706.
- Shimizu T, Nakano T, Takamizawa D, Desaki Y, Ishii-Minami N, Nishizawa Y, et al. Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. *Plant J*. 2010;64:204–14.
- Vaid N, Pandey PK, Tuteja N. Genome-wide analysis of lectin receptor-like kinase family from Arabidopsis and rice. *Plant Mol Biol*. 2012;80:365–88.
- Vaid N, Macovei A, Tuteja N. Knights in action: lectin receptor-like kinases in plant development and stress responses. *Mol Plant*. 2013;6:1405–18.
- Wang Y, Weide R, Govers F, Bouwmeester K. L-type lectin receptor kinases in *Nicotiana benthamiana* and tomato and their role in *Phytophthora* resistance. *J Exp Bot*. 2015;66:6731–43.
- Shumayla SS, Pandey AK, Singh K, Upadhyay SK. Molecular characterization and global expression analysis of lectin receptor kinases in bread wheat (*Triticum aestivum*). *PLoS One*. 2016;11:e0153925.
- Balagué C, Gouget A, Bouchez O, Souriac C, Haget N, Boutet-Mercey S, et al. The Arabidopsis thaliana lectin receptor kinase LecRK-I.9 is required for full resistance to *Pseudomonas syringae* and affects jasmonate signalling. *Mol Plant Pathol*. 2017;18:937–48.
- Wang Y, Cordewener JHG, America AHP, Shan W, Bouwmeester K, Govers F. Arabidopsis lectin receptor kinases LecRK-IX.1 and LecRK-IX.2 are functional analogs in regulating Phytophthora resistance and plant cell death. *Mol Plant-Microbe Interact*. 2015;28:1032–48.
- Bouwmeester K, Govers F. Arabidopsis L-type lectin receptor kinases: phylogeny, classification, and expression profiles. *J Exp Bot*. 2009;60:4383–96.
- Huang P-Y, Yeh Y-H, Liu A-C, Cheng C-P, Zimmerli L. The Arabidopsis LecRK-VI.2 associates with the pattern-recognition receptor FLS2 and primes *Nicotiana benthamiana* pattern-triggered immunity. *Plant J*. 2014;79:243–55.
- Choi J, Tanaka K, Cao Y, Qi Y, Qiu J, Liang Y, et al. Identification of a plant receptor for extracellular ATP. *Science*. 2014;343:290–4.
- Cao Y, Tanaka K, Nguyen CT, Stacey G. Extracellular ATP is a central signaling molecule in plant stress responses. *Curr Opin Plant Biol*. 2014;20:82–7.
- Van Damme EJM, Nakamura-Tsuruta S, Smith DF, Ongenaert M, Winter HC, Rougé P, et al. Phylogenetic and specificity studies of two-domain GNA-related lectins: generation of multispecificity through domain duplication and divergent evolution. *Biochem J*. 2007;404:51–61.
- Van Damme EJM, Allen AK, Peumans WJ. Isolation and characterization of a lectin with exclusive specificity towards mannose from snowdrop (*Galanthus nivalis*) bulbs. *FEBS Lett*. 1987;215:140–4.
- Hester G, Kaku H, Goldstein IJ, Wright CS. Structure of mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. *Nat Struct Biol*. 1995;2:472–9.
- Damme EJM, Allen AK, Peumans WJ. Related mannose-specific lectins from different species of the family Amaryllidaceae. *Physiol Plant*. 1988;73:52–7.
- Van Damme EJM, Goldstein IJ, Peumans WJ. A comparative study of mannose-binding lectins from the amaryllidaceae and alliaceae. *Phytochemistry*. 1991;30:509–14.
- Van Damme EJM, Goldstein IJ, Vercammen G, Vuylsteke J, Peumans WJ. Lectins of members of the Amaryllidaceae are encoded by multigene families which show extensive homology. *Physiol Plant*. 1992;86:245–52.
- Peumans WJ, Barre A, Bras J, Rougé P, Proost P, Van Damme EJM. The liverwort contains a lectin that is structurally and evolutionarily related to the monocot mannose-binding lectins. *Plant Physiol*. 2002;129:1054–65.
- Kai G, Zhao L, Zheng J, Zhang L, Miao Z, Sun X, et al. Isolation and characterization of a new mannose-binding lectin gene from *Taxus media*. *J Biosci*. 2004;29:399–407.

46. Sherman-Broyles S, Boggs N, Farkas A, Liu P, Vrebalov J, Nasrallah ME, et al. S locus genes and the evolution of self-fertility in *Arabidopsis thaliana*. *Plant Cell*. 2007;19:94–106.
47. Kusaba M, Dwyer K, Hendershot J, Vrebalov J, Nasrallah JB, Nasrallah ME. Self-incompatibility in the genus *Arabidopsis*: characterization of the S locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. *Plant Cell*. 2001;13:627–43.
48. Shiu SH, Bleecker AB. Plant receptor-like kinase gene family: diversity, function, and signaling. *Sci STKE*. 2001;2001:re22.
49. Tordai H, Bányai L, Patthy L. The PAN module: the N-terminal domains of plasminogen and hepatocyte growth factor are homologous with the apple domains of the prekallikrein family and with a novel domain found in numerous nematode proteins. *FEBS Lett*. 1999;461:63–7.
50. Shearer LA, Anderson LK, de Jong H, Smit S, Goicoechea JL, Roe BA, et al. Fluorescence *in situ* hybridization and optical mapping to correct scaffold arrangement in the tomato genome. *G3*. 2014;4:1395–405.
51. Mitchell A, Chang H-Y, Daugherty L, Fraser M, Hunter S, Lopez R, et al. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res*. 2015;43:D213–21.
52. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–7.
53. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci*. 1992;8:275–82.
54. Lefort V, Longueville J-E, Gascuel O. SMS: Smart Model Selection in PhyML. *Mol Biol Evol*. 2017;34:2422–4.
55. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30:1312–3.
56. Stöver BC, Müller KF. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics*. 2010;11:7.
57. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37:W202–8.
58. Tanz SK, Castleden I, Hooper CM, Vacher M, Small I, Millar HA. SUBA3: a database for integrating experimentation and prediction to define the SUBcellular location of proteins in *Arabidopsis*. *Nucleic Acids Res*. 2013;41: D1185–91.
59. Hooper CM, Tanz SK, Castleden IR, Vacher MA, Small ID, Millar AH. SUBAcon: a consensus algorithm for unifying the subcellular localization data of the *Arabidopsis* proteome. *Bioinformatics*. 2014;30:3356–64.
60. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc*. 2007;2:953–71.
61. Yu C-S, Chen Y-C, Lu C-H, Hwang J-K. Prediction of protein subcellular localization. *Proteins*. 2006;64:643–51.
62. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–402.
63. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res*. 2015;43: D222–6.
64. Guo Y-L, Fitz J, Schneeberger K, Ossowski S, Cao J, Weigel D. Genome-wide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in *Arabidopsis*. *Plant Physiol*. 2011;157:757–69.
65. Kim HS, Jung MS, Lee SM, Kim KE, Byun H, Choi MS, et al. An S-locus receptor-like kinase plays a role as a negative regulator in plant defense responses. *Biochem Biophys Res Commun*. 2009;381:424–8.
66. Ranf S, Gisch N, Schäffer M, Illig T, Westphal L, Knirel YA, et al. A lectin S-domain receptor kinase mediates lipopolysaccharide sensing in *Arabidopsis thaliana*. *Nat Immunol*. 2015;16:426–33.
67. ten Hove CA, Bocharovits Z, Jansweijer VMA, Koning FG, Berke L, Sanchez-Perez GF, et al. Probing the roles of LRR RLK genes in *Arabidopsis thaliana* roots using a custom T-DNA insertion set. *Plant Mol Biol*. 2011;76:69–83.
68. Swain S, Singh N, Nandi AK. Identification of plant defence regulators through transcriptional profiling of *Arabidopsis thaliana cdd1* mutant. *J Biosci*. 2015;40:137–46.
69. Ko J-H, Kim JH, Jayanty SS, Howe GA, Han K-H. Loss of function of COBRA, a determinant of oriented cell expansion, invokes cellular defence responses in *Arabidopsis thaliana*. *J Exp Bot*. 2006;57:2923–36.
70. Trontin C, Kiani S, Corwin JA, Hématy K, Yansouni J, Klebenstein DJ, et al. A pair of receptor-like kinases is responsible for natural variation in shoot growth response to mannitol treatment in *Arabidopsis thaliana*. *Plant J*. 2014;78:121–33.
71. Lan P, Li W, Schmidt W. Genome-wide co-expression analysis predicts protein kinases as important regulators of phosphate deficiency-induced root hair remodeling in *Arabidopsis*. *BMC Genomics*. 2013;14:210.
72. Chae L, Sudat S, Dudoit S, Zhu T, Luan S. Diverse transcriptional programs associated with environmental stress and hormones in the *Arabidopsis* receptor-like kinase gene family. *Mol Plant*. 2009;2:84–107.
73. Kumari S, Shridhar S, Singh D, Priya P, Farmer R, Hundal J, et al. The role of lectins and HD-ZIP transcription factors in isoprenoid based plant stress responses. *Proc Indian Natl Sci Acad*. 2012;78:671–91.
74. Kalladan R, Lasky JR, Chang TZ, Sharma S, Juenger TE, Verslues PE. Natural variation identifies genes affecting drought-induced abscisic acid accumulation in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A*. 2017;114: 11536–41.
75. Qutob D, Kemmerling B, Brunner F, Küfner I, Engelhardt S, Gust AA, et al. Phytotoxicity and innate immune responses induced by Nep1-Like proteins. *Plant Cell*. 2006;18:3721.
76. Tischner R, Koltermann M, Hesse H, Plath M. Early responses of *Arabidopsis thaliana* to infection by *Verticillium longisporum*. *Physiol Mol Plant Pathol*. 2010;74:419–27.
77. Ma S, Gong Q, Bohnert HJ. Dissecting salt stress pathways. *J Exp Bot*. 2006; 57:1097–107.
78. Zhu Q-H, Stephen S, Kazan K, Jin G, Fan L, Taylor J, et al. Characterization of the defense transcriptome responsive to *Fusarium oxysporum*-infection in *Arabidopsis* using RNA-seq. *Gene*. 2013;512:259–66.
79. Deb S, Sankaranarayanan S, Wewala G, Widdup E, Samuel MA. The S-Domain receptor kinase *Arabidopsis* Receptor Kinase2 and the U Box/ Armadillo repeat-containing E3 Ubiquitin Ligase9 module mediates lateral root development under phosphate starvation in *Arabidopsis*. *Plant Physiol*. 2014;165:1647–56.
80. Mishiba K-I, Nagashima Y, Suzuki E, Hayashi N, Ogata Y, Shimada Y, et al. Defects in IRE1 enhance cell death and fail to degrade mRNAs encoding secretory pathway proteins in the *Arabidopsis* unfolded protein response. *Proc Natl Acad Sci U S A*. 2013;110:5713–8.
81. Samuel MA, Mudgil Y, Salt JN, Delmas F, Ramachandran S, Chilelli A, et al. Interactions between the S-domain receptor kinases and AtPUB-ARM E3 ubiquitin ligases suggest a conserved signaling pathway in *Arabidopsis*. *Plant Physiol*. 2008;147:2084–95.
82. Birkenbihl RP, Kracher B, Somssich IE. Induced genome-wide binding of three *Arabidopsis* WRKY transcription factors during early MAMP-triggered immunity. *Plant Cell*. 2017;29:20–38.
83. Schluepmann H, van Dijken A, Aghdasi M, Wobbes B, Paul M, Smeeckens S. Trehalose mediated growth inhibition of *Arabidopsis* seedlings is due to trehalose-6-phosphate accumulation. *Plant Physiol*. 2004;135:879–90.
84. New S-A, Piater LA, Dubery IA. In silico characterization and expression analysis of selected *Arabidopsis* receptor-like kinase genes responsive to different MAMP inducers. *Biol Plant*. 2015;59:18–28.
85. Blaum BS, Mazzotta S, Nöldeke ER, Halter T, Madlung J, Kemmerling B, et al. Structure of the pseudokinase domain of BIR2, a regulator of BAK1-mediated immune signaling in *Arabidopsis*. *J Struct Biol*. 2014;186:112–21.
86. Chen L-J, Wuriyangan H, Zhang Y-Q, Duan K-X, Chen H-W, Li Q-T, et al. An S-domain receptor-like kinase, OsSIK2, confers abiotic stress tolerance and delays dark-induced leaf senescence in rice. *Plant Physiol*. 2013;163:1752–65.
87. Amil-Ruiz F. Molecular mechanisms of strawberry plant defence against *Colletotrichum acutatum* Caballero Repullo JL, Muñoz Blanco J, editors. Universidad de Córdoba; 2013.
88. Sanabria NM, van Heerden H, Dubery IA. Molecular characterisation and regulation of a *Nicotiana tabacum* S-domain receptor-like kinase gene induced during an early rapid response to lipopolysaccharides. *Gene*. 2012; 501:39–48.
89. Chen X, Shang J, Chen D, Lei C, Zou Y, Zhai W, et al. AB-lectin receptor kinase gene conferring rice blast resistance. *Plant J*. 2006;46:794–804.
90. Fuller VL, Lilley CJ, Atkinson HJ, Urwin PE. Differential gene expression in *Arabidopsis* following infection by plant-parasitic nematodes *Meloidogyne incognita* and *Heterodera schachtii*. *Mol Plant Pathol*. 2007;8: 595–609.
91. Sun X, Sun M, Luo X, Ding X, Ji W, Cai H, et al. A Glycine soja ABA-responsive receptor-like cytoplasmic kinase, GsRLCK, positively controls plant tolerance to salt and drought stresses. *Planta*. 2013;237:1527–45.
92. Hanks SK, Hunter T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J*. 1995;9:576–96.

93. Shiu SH, Bleecker AB. Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *Proc Natl Acad Sci U S A*. 2001;98:10763–8.
94. Kim J-G, Li X, Roden JA, Taylor KW, Aakre CD, Su B, et al. *Xanthomonas* T3S effector XopN suppresses PAMP-triggered immunity and interacts with a tomato atypical receptor-like kinase and TFT1. *Plant Cell*. 2009;21:1305–23.
95. Niu D, Lii YE, Chellappan P, Lei L, Peralta K, Jiang C, et al. miRNA863-3p sequentially targets negative immune regulator ARLPKs and positive regulator SERRATE upon bacterial infection. *Nat Commun*. 2016;7:11324.
96. Liang X, Shen K, Lichtenberg J, Wyatt SE, Welch LR. An integrated bioinformatics approach to the discovery of cis-regulatory elements involved in plant gravitropic signal transduction. *Inter J Comp Biosci*. 2010;1:33–54.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

